



Mapping the genome of *Plasmodium falciparum* on the drug-like chemical space reveals novel anti-malarial targets and potential drug leads

Jensen, Kasper; Plichta, Damian Rafal; Panagiotou, Gianni; Kouskoumvekaki, Irene

Published in:
Molecular BioSystems

Link to article, DOI:
[10.1039/c2mb00008c](https://doi.org/10.1039/c2mb00008c)

Publication date:
2012

Document Version
Publisher's PDF, also known as Version of record

[Link back to DTU Orbit](#)

Citation (APA):
Jensen, K., Plichta, D. R., Panagiotou, G., & Kouskoumvekaki, I. (2012). Mapping the genome of *Plasmodium falciparum* on the drug-like chemical space reveals novel anti-malarial targets and potential drug leads. *Molecular BioSystems*, 8(6), 1678-1685. <https://doi.org/10.1039/c2mb00008c>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Cite this: *Mol. BioSyst.*, 2012, **8**, 1678–1685

www.rsc.org/molecularbiosystems

PAPER

Mapping the genome of *Plasmodium falciparum* on the drug-like chemical space reveals novel anti-malarial targets and potential drug leads†

Kasper Jensen, Damian Plichta, Gianni Panagiotou* and Irene Kouskoumvekaki*

Received 6th January 2012, Accepted 28th February 2012

DOI: 10.1039/c2mb00008c

The parasite *Plasmodium falciparum* is the main agent responsible for malaria. In this study, we exploited a recently published chemical library from GlaxoSmithKline (GSK) that had previously been confirmed to inhibit parasite growth of the wild type (3D7) and the multi-drug resistance (D2d) strains, in order to uncover the weak links in the proteome of the parasite. We predicted 293 proteins of *P. falciparum*, including the six out of the seven verified targets for *P. falciparum* malaria treatment, as targets of 4645 GSK active compounds. Furthermore, we prioritized druggable targets, based on a number of factors, such as essentiality for growth, lack of homology with human proteins, and availability of experimental data on ligand activity with a non-human homologue of a parasite protein. We have additionally prioritized predicted ligands based on their polypharmacology profile, with focus on validated essential proteins and the effect of their perturbations on the metabolic network of *P. falciparum*, as well as indication of drug resistance emergence. Finally, we predict potential off-target effects on the human host with associations to cancer, neurological and dermatological disorders, based on integration of available chemical–protein and protein–protein interaction data. Our work suggests that a large number of the *P. falciparum* proteome is potentially druggable and could therefore serve as novel drug targets in the fight against malaria. At the same time, prioritized compounds from the GSK library could serve as lead compounds to medicinal chemists for further optimization.

Introduction

Malaria is one of the most hazardous tropical parasitic diseases in humans and can be dated back to 2800 BC.^{1,2} Fifty years ago, malaria had been almost eliminated from many areas of the world through a combination of long acting pharmaceuticals and vector control interventions.^{3,4} However, the global spread of drug resistance to all but one of the five major classes of anti-malarial drugs,^{5,6} together with a collapse of vector control programs, resulted in resurgence in disease morbidity and mortality by the 1980s. About 50% of the world's population live in endemic areas, situated mostly in Africa, Asia, and South America.⁷ With approximately 250 million cases and 850 000 attributed deaths reported globally in 2009,⁷ malaria is one of the most severe infectious diseases primarily affecting the world's most disadvantaged populations and killing more under-5-year olds than any other infectious agent. Up to now no approved vaccine is available, while researchers in Cambodia have reported the first signs of

resistance to artemisinins, the essential ingredient of existing malarial treatments for 100 million patients annually.⁸ The grim malaria statistics could rise even higher if resistance to the existing anti-malarial drugs develops further.⁹

Malaria is generally caused by four apicomplexan parasites of the genus *Plasmodium*, including *Plasmodium falciparum* (most lethal), *P. vivax*, *P. ovale*, and *P. malariae*. The malaria parasite *P. falciparum* exhibits a complex life cycle that involves infection of a vertebrate human host by the female *Anopheles* mosquito during a blood meal. Unlike the human host, *P. falciparum* lacks the ability to salvage pyrimidine bases and thus is entirely dependent on *de novo* biosynthesis.^{10,11} Clinical manifestations begin when parasites enter erythrocytes. The erythrocytic cycle of *P. falciparum* involves a massive amplification of the parasite population through periodic cycles of invasion, growth, division and egress from erythrocytes.

In 2002, the genome of *P. falciparum* was sequenced,¹² renewing hope that progress towards reducing the burden of malaria would be greatly accelerated by understanding the biology of the parasite at the molecular level. Rational drug design was introduced in malaria research at the same time, with great optimism that new drugs would emerge based on the knowledge of the biological target. Although our understanding of the parasite's biology has significantly increased in

Department of Systems Biology, Technical University of Denmark, Kgs Lyngby 2800, Denmark. E-mail: gpa@bio.dtu.dk, irene@cbs.dtu.dk

† Electronic supplementary information (ESI) available. See DOI: 10.1039/c2mb00008c

the last decade, no new class of antimalarials has been introduced into clinical practice since 1996.¹⁶ Approximately half of the predicted proteins to date are of uncharacterized function, with little sequence similarity outside the *Plasmodium* genus, while a large number of genes and gene families are unique to *P. falciparum*.^{12,13} Amongst the ~5300 proteins encoded in *P. falciparum* genome, only seven, namely PF10_0084 (tubulin beta chain), PF10_0289 (adenosine diaminase), PFD0830w (bifunctional dihydrofolate reductase-thymidylate synthase), PFE0520c (topoisomerase I), PFF0160c (dihydroorotate dehydrogenase), PF10_0334 (flavoprotein subunit of succinate dehydrogenase) and PFI1020c (inosine-5'-monophosphate dehydrogenase), are currently verified drug targets.¹⁴

Clinical experience has shown that drugs targeting a single molecular pathway, monotherapy or highly mutable drug targets, have each led to resistance and are undesirable in effective long-term strategies against multi-drug resistant malaria.¹⁵ Parasite-based screening is an alternative approach to rational drug design with the advantage of identifying compounds that may act on more than one molecular target or pathway. In such compound screening endeavor that was recently published, Gamo *et al.*¹⁶ used the GlaxoSmithKline (GSK) in-house chemical library to evaluate almost 2 million compounds against asexual blood stage *P. falciparum*. Setting the threshold for growth inhibition to greater than 80%, at a concentration cut-off of 2 micromolar, the authors identified more than 13 500 active compounds, 8000 of these being equally active against multi-drug-resistant *P. falciparum* parasites.

Gamo and collaborators, however, did not identify the parasite drug targets, which is the specific aim of the present study. For anti-infectious drug discovery, where pharmaceutically investigated targets only represent a small portion of the whole pathogen's genome, it is more challenging to establish a proteome-wide drug-target network.¹⁷ Here we use the wealth of information from the GSK compound collection to link the proteome of *P. falciparum* with the bioactive chemical space, following a systems chemical biology approach fueled by similarity-based virtual screening and protein homologue analysis. One of our primal aims is to identify druggable proteins in *P. falciparum* and to prioritize them on the basis of essentiality for growth, lack of homology with human proteins and availability of experimental data that support prediction of binding to an active GSK compound. At the same time, we aim to identify compounds from the GSK collection that are capable of escaping drug resistance, have polypharmacology profiles that cover multiple essential targets and perturb different points of the parasite's metabolic network. Our results could also be seen in the light of drug synergy, and assist in the selection of compounds with synergistic effects that could be used in a combinatorial antimalarial therapy. Finally, we evaluate *in silico* potential toxic and off-target interactions of the predicted polyactive GSK compounds with the human proteome, by integrating available, experimental information from chemical-protein and protein-protein interaction databases. Our study is the first to follow up on the published GSK library of compounds and contributes to the selection of durable anti-malarial drug targets and chemical molecules for further preclinical evaluation.

Results and discussion

The druggable landscape of the *P. falciparum* proteome

The chemical library of compounds with anti-malaria properties from the GSK in-house collection is the largest published to date, but it lacks information regarding interactions between the tested compounds and the parasite's proteome. Since the anti-malarial properties of the 13 533 GSK compounds are based on an *in vivo* screen, we initially ventured a concerted systems chemical biology approach to create an interaction map of the active compounds from GSK on the parasite's protein space. ChemProt,¹⁸ a database compiled *in-house* from multiple chemical-protein annotation resources, with more than 700 000 unique chemicals and biological annotation for 30 578 targets, was surveyed in two ways: firstly, for direct experimental bioactivity information concerning the 13 533 GSK compounds, and secondly, for predicted bioactivities *via* structural similarities with compounds from ChemProt, on the premise that structurally similar molecules are expected to exhibit similar biological activities.

Not surprisingly, considering the uniqueness of this dataset, there are only 205 GSK compounds with recorded experimental activity in ChemProt, involving 86 protein targets, 51 of which have a homologue in *P. falciparum*. The target space of the GSK compounds was subsequently enriched with 250 protein targets in the proteome of *P. falciparum* associated with 4495 GSK compounds that have Tanimoto coefficient similarity of at least 0.85 with a bioactive compound in ChemProt. It has been previously reported in the scientific literature that compounds with structural similarity above this threshold tend to exert similar biological activities.¹⁹ The full network of GSK compounds and the druggable *P. falciparum* proteome consists of 4645 unique small molecules and 293 protein targets. The genomic location of the 293 proteins was identified using PlasmoDB (www.plasmodb.org/) (Fig. 1A). The genomic location of the genes coding for the 293 protein targets, taking into consideration the size of the chromosome, indicates with 95% confidence a preference for chromosomes 3, 13 and 14 while lower preference is indicated for chromosomes 4, 7, 8, 9 (Fig. 1B). To conceptualize the sensitivity of *P. falciparum* to the GSK compounds, we assigned GO-terms to the 293 protein targets using Blast2GO (www.blast2go.org/). The molecular function annotation revealed that the majority of the protein targets are involved in nucleotide binding, followed by protein binding, transporter activity, transcriptional regulation and protein kinase activity, while their involvement into biological processes includes amino-acid and lipid metabolism, protein modification and translation (see File S1, ESI[†]).

Protein kinases are considered as particularly attractive drug-target classes, not only because it is well established that small molecules can bind to their catalytic cleft but also because protein phosphorylation is one of the most important and pleiotropic modes of regulation of cell physiology.²⁰ Lipid metabolism, which is almost nonfunctional in the uninfected erythrocyte, is dramatically elevated during the intraerythrocytic development of the malaria parasite contributing mainly to membrane biogenesis. The unique features in lipid metabolism of *P. falciparum* are ideal for malaria chemotherapeutic purposes.²¹ Transporters, providing essential nutrients and

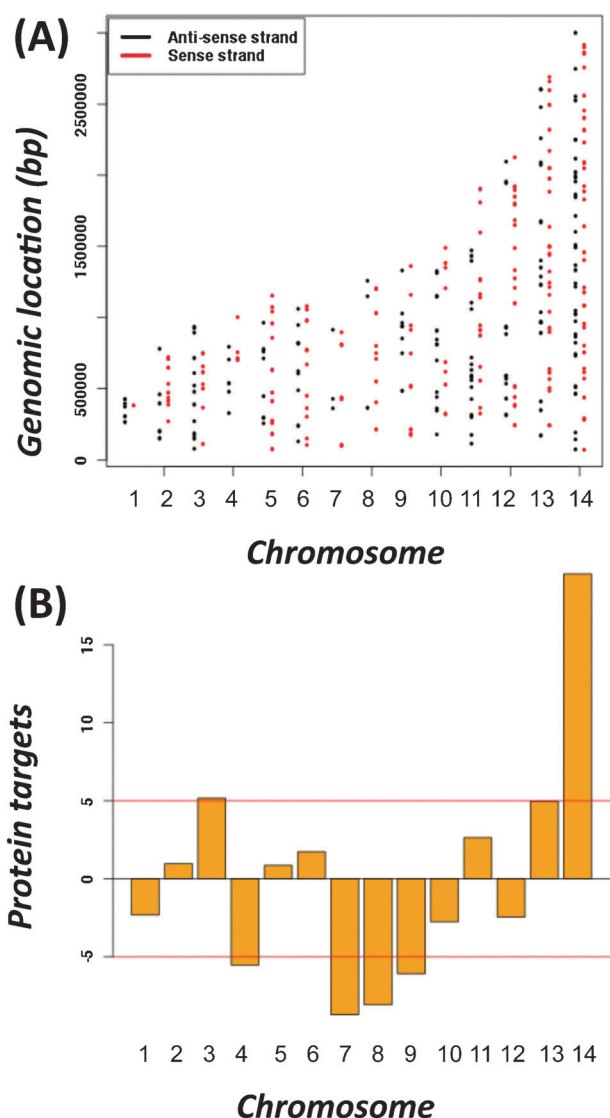


Fig. 1 Predicted protein targets of the GlaxoSmithKline compounds in the genome of *P. falciparum*. (A) Genomic location of predicted targets of GSK compounds. The x-axis indicates the chromosomes 1 to 14 and the y-axis indicates the location of coding genes in base pairs (bp). Red: proteins coded at the sense-strand; black: proteins coded at the anti-sense strand. (B) Target preference per chromosome. The y-axis shows the difference between the predicted and expected number of protein targets on each chromosome (corrected for the size of the chromosome). The bar indicates 95% confidence threshold, using Student's *t*-test.

disposing toxic wastes as well as being components of important regulatory pathways in *P. falciparum*, are excellent targets. Transporters can provide delivery routes for drugs that target essential parasite mechanisms but are also involved in mediating drug resistance.^{22,23}

A detailed matrix of all predicted interactions between GSK compounds and the *P. falciparum* proteome is provided in File S2 (ESI†). It is worth noting that six out of the seven verified targets for *P. falciparum* malaria treatment, namely PF10_0084, PF10_0289, PFD0830w, PFE0520c, PFF0160c, and PFI1020c, are predicted to interact with compounds from the GSK library. 2133 GSK compounds (almost half of the total number of predictions) selectively target one amongst a

subset of 130 proteins of *P. falciparum*. This finding could lead to the assumption that these 130 proteins are important for the growth of the parasite, while the fact that five of the verified malaria targets are included in this list reinforces this statement. 77 of these do not have a human homologue in NCBI HomoloGene,²⁴ which is another desirable feature when selecting parasitic targets for drug development.

We acknowledge here that our ability to identify protein targets in the proteome of *P. falciparum* that may be important for the growth of the parasite is limited by the chemical and target space covered by the ChemProt database. If a *P. falciparum* protein does not carry adequate sequence similarity with any of the proteins that are included in the repository, we will be, of course, unable to identify it as a potential drug target. Similarly, if a GSK compound does not have adequate structural similarity with any of the ligands in ChemProt, we fail to capture its protein target(s) in the *P. falciparum* proteome. Nevertheless, by using the present systems chemical biology approach we manage to move away from conventional methods that stay in the space of the known antimalarial drugs and their known targets. To our knowledge, this is the first drug–target network that covers this many proteins in the *P. falciparum* proteome and all compounds that have shown antimalarial activity on *in vitro* parasite screens.

Design of multi-target therapies against malaria

Multi-target therapy is believed to be more effective than single-target therapy when treating infectious diseases.¹⁷ A promising drug candidate should, therefore, not be selective to a single target, but rather be polyactive, *i.e.* have an activity profile that includes more than one protein essential for the parasite—preferably with no human homologues—in order to avoid drug resistance and human toxicity. When one of the targets is already a known antimalarial drug target, there are increased chances that the given compound has an effective polyactivity profile. Further to this, prior experimental evidence for biological activity of the given compound against a homologue of a *P. falciparum* protein increases the confidence of the predicted interaction between the two. We have, thus, analyzed all predicted interactions taking all the above considerations into account and in the heatmap of Fig. 2 we zoom into the drug–target space of polyactive GSK compounds with those *P. falciparum* proteins that fulfill these requirements.

For example, GSK compound hispidulin (CID 5281628) is predicted to bind to three known antimalarial drug targets, namely PF10_0084, PFE0520c and PFD0830w, the latter of which has no known human homologue. Hispidulin is additionally predicted to bind to 28 other proteins of *P. falciparum* that may also be important for the growth of the parasite, 15 of which have no known human homologues. Hispidulin is a small flavonoid molecule from *Artemisia vestita*, a traditional Tibetan medicinal plant, which is known for its anti-inflammatory, antifungal, anti-proliferative, anti-thrombosis and anti-oxidative activities.²⁵ Recently, hispidulin has been studied and found active on suppressing angiogenesis and pancreatic tumor growth,²⁵ while its activity on the *P. falciparum* screen conducted by GSK makes it a very promising lead compound for further evaluation and optimization.

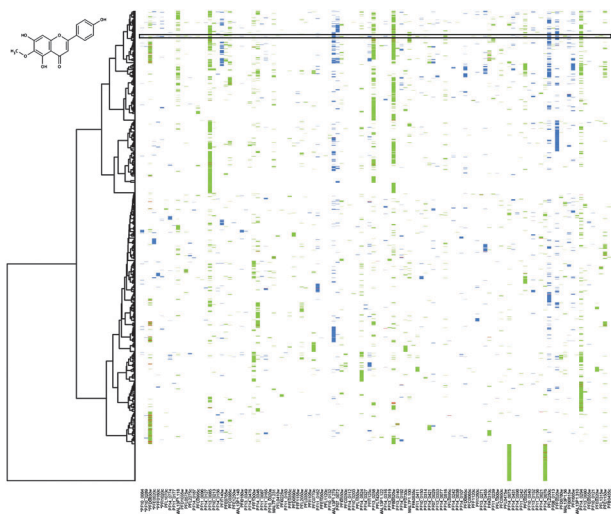


Fig. 2 Heatmap zoomed on the predicted interactions between polyactive GSK compounds and prioritized *P. falciparum* proteins. Green: protein target with no human homologue. Red: known experimental activity with a non-human protein homologue of *P. falciparum*. Blue: rest of predicted interactions within the subset of the 130 prioritized proteins. The first six targets on the left side of the x-axis are the known antimalarial drug targets. On the top left, the structure of GSK compound hispidulin (CID 5281628), whose interactions are included in the rectangle. The raw data behind the heatmap are provided in File S3 (ESI†).

On a subsequent analysis of the predicted drug–target interactions, we used the recently compiled list of 30 experimentally

validated gene knockouts and phenotypes in *P. falciparum* by Plata *et al.*²⁶ This list of genes has resulted from targeted inhibitions of enzymatic activities from drugs, and fifteen of them are also included in our list of 293 predicted targets. Fig. 3 zooms into the results within this area of the druggable proteome space of *P. falciparum*, where polyactive compounds are predicted to target the experimentally validated essential genes. Compounds in bold have shown *in vitro* activity against both the wild type *P. falciparum* 3D7 and the D2d resistant strain.¹⁶ For the majority of the fifteen essential genes, there are GSK compounds that contain more than one of them in their activity profile. Multiple polymorphisms in the genes coding for antimalarial protein targets have been associated with emergence of drug resistance, thus, having an activity profile that includes more than one essential protein, could be an efficient way for the drug to escape resistance. Only in the case of PF14_0053 (ribonucleotide reductase), PF14_0425 (fructose-bisphosphate aldolase) and PFB0295w (adenylosuccinate lyase) the polyactive compounds that are predicted to bind to them do not have any other of the experimentally validated essential proteins in their activity profile. Surprisingly, most of these compounds, *e.g.* all the compounds targeting PF14_0053, have shown *in vitro* activity against both the wild type and the resistant strain. Similarly, most of the compounds from the cluster that is associated with the two known drug targets PF10_0289 and PFD0830w seem to have the right polyactivity profile that manages to evade drug resistance. One of them, CID 23618032, has been annotated in ChemProt with experimental binding activity of 6.41 log

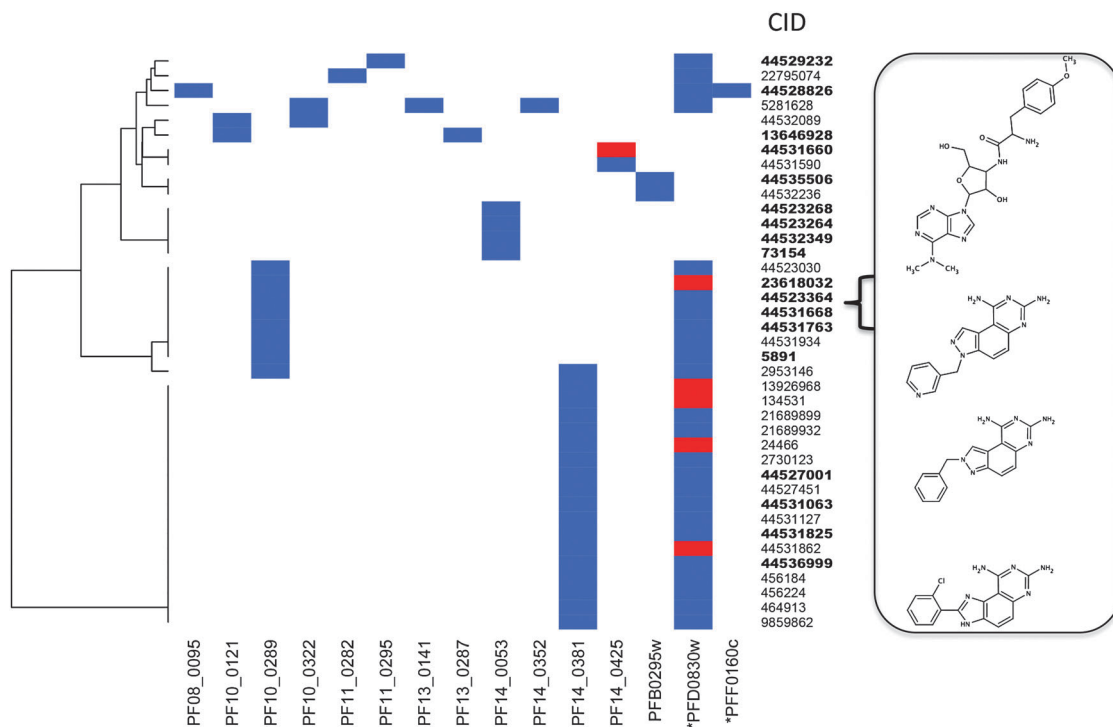


Fig. 3 Heatmap with predicted interactions between GSK compounds and *P. falciparum* experimentally validated essential proteins. Red: validated experimental activity with a non-human protein homologue of *P. falciparum*. Blue: rest of predicted interactions. GSK compounds are named according to PubChem CID. Protein names with asterisk are known antimalarial drug targets. Compounds in bold are active on both the wild type and resistant strains. The structures of such compounds that are predicted to bind to two known drug targets are shown in the box on the right. With the exception of PF14_0053, PF14_0425 and PFB0295w, only compounds that target at least two essential proteins have been included.

(1/C) with a close homologue of PFD0830w, namely dihydrofolate reductase from *Mus musculus*, which gives support to the prediction. For these promising drug leads, one should look deeper into their full activity profile, as activity on non-essential proteins should not be underestimated, because of the emergence of “synthetic lethality”, where genes that are not essential individually are shown to be essential in combination.

In addition to the above analyses, we used the recently reconstructed genome-scale flux balanced metabolic model of *P. falciparum*²⁶ to identify the perturbations caused in the metabolism by the GSK compounds from a network level perspective. In Fig. 4, we focus on the essential proteins from Plata *et al.*²⁶ and the 39 promising GSK compounds shown in Fig. 3, advocating that disturbing the robustness of the metabolic network over multiple biological pathways will increase drug effectiveness and prevent the development of drug resistance. There is a cluster of compounds in the heatmap of Fig. 3 that are active on both wild type and resistant strains and are predicted to bind to both known antimalarial drug targets PF10_0289 and PFD0830w. Projecting these two targets on the metabolic network of Fig. 4, we see that they are involved in different pathways, namely nucleotide metabolism and biosynthesis of other secondary metabolites. It would, thus, be of interest to look deeper into the specific metabolic pathways and investigate whether and how a simultaneous perturbation of both could lead to parasite elimination and escape of drug resistance.

Off-target interactions with the human proteome

We further studied potentially toxic off-target interactions with human proteins and the 2512 polyactive GSK compounds, extracting information from the Comparative Toxicogenomics Database,²⁷ one of the databases within ChemProt, where drugs and environmental compounds are associated with toxicologically important proteins. CDK has information for only five GSK compounds that have been previously associated in the literature with adverse and toxic effects. Tryptanthrine (CID 73549) is reported to result in decreased expression of ABCB1 and TP53 that are involved in different types of cancer. Olvanil (CID 5311093) has been linked to TRPV1 oligomer formation and increased expression of TAC1, both of which are related to a number of esophageal and CNS diseases. Trimethoprim (CID 5578), a known antibacterial and currently in use as antimalarial medication, has been associated with rhabdomyolysis, as well as with antitumor activity in transitional cell carcinoma and urinary bladder neoplasms, when administered together with Sulfamethoxazole. Cycloheximide (CID 6197) has been reported to interact with 300 different human proteins associated with a multitude of diseases, where malignant neoplasms are ranked with the highest inference scores. Finally, pyrimethamine (CID 4993), which is one of the safest drugs used in malaria treatment, has been associated with decreased activity of GSTP1 and CYP2C9 proteins and increased expression of FOS and DDIT3 mRNA. FOS and GSTP1 are linked to drug hypersensitivity, which is a well-known side effect of the drug.²⁸

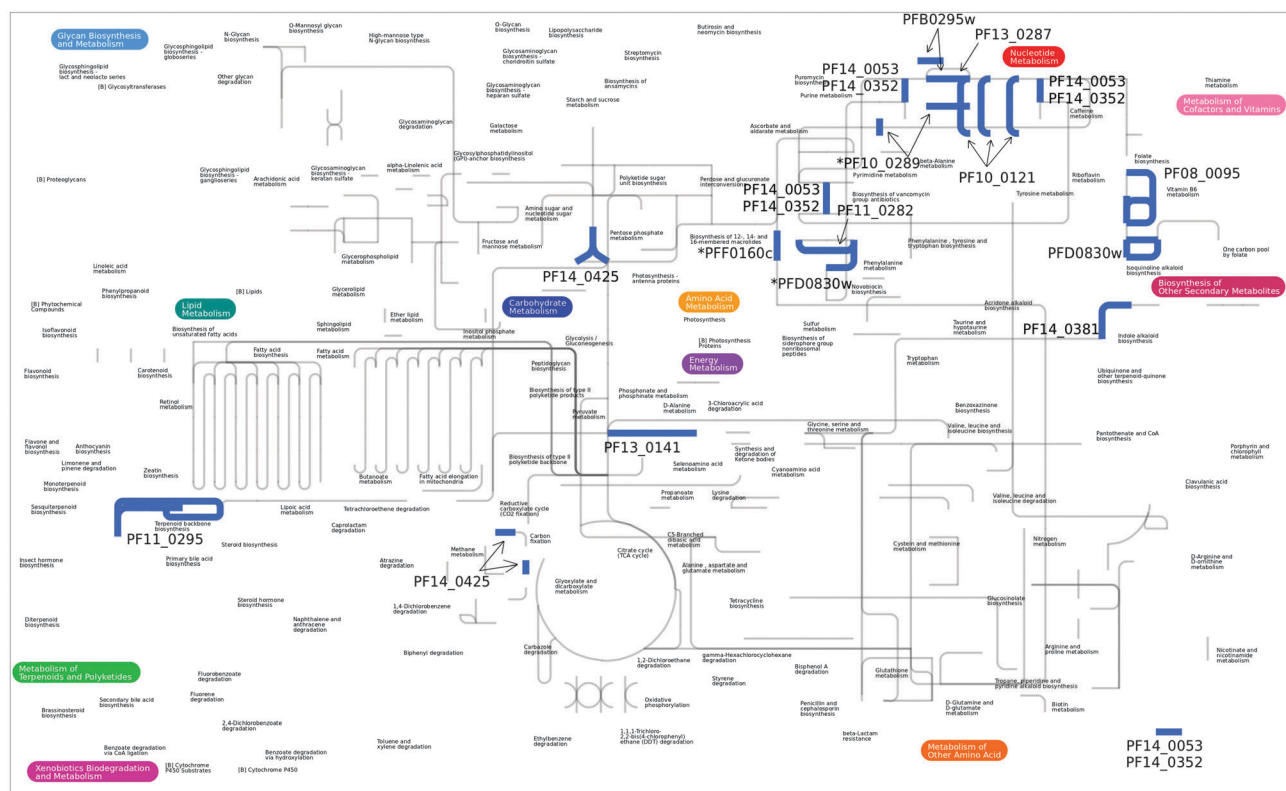


Fig. 4 Visualization of the metabolic pathways of *P. falciparum*, where the 15 essential genes targeted by GSK compounds are participating. The metabolic map was created from *P. falciparum* 3D7, species code: 36329. PF10_0322 has been omitted, as it could not be associated with a KEGG ID. Protein names with asterisk are known antimalarial drug targets.

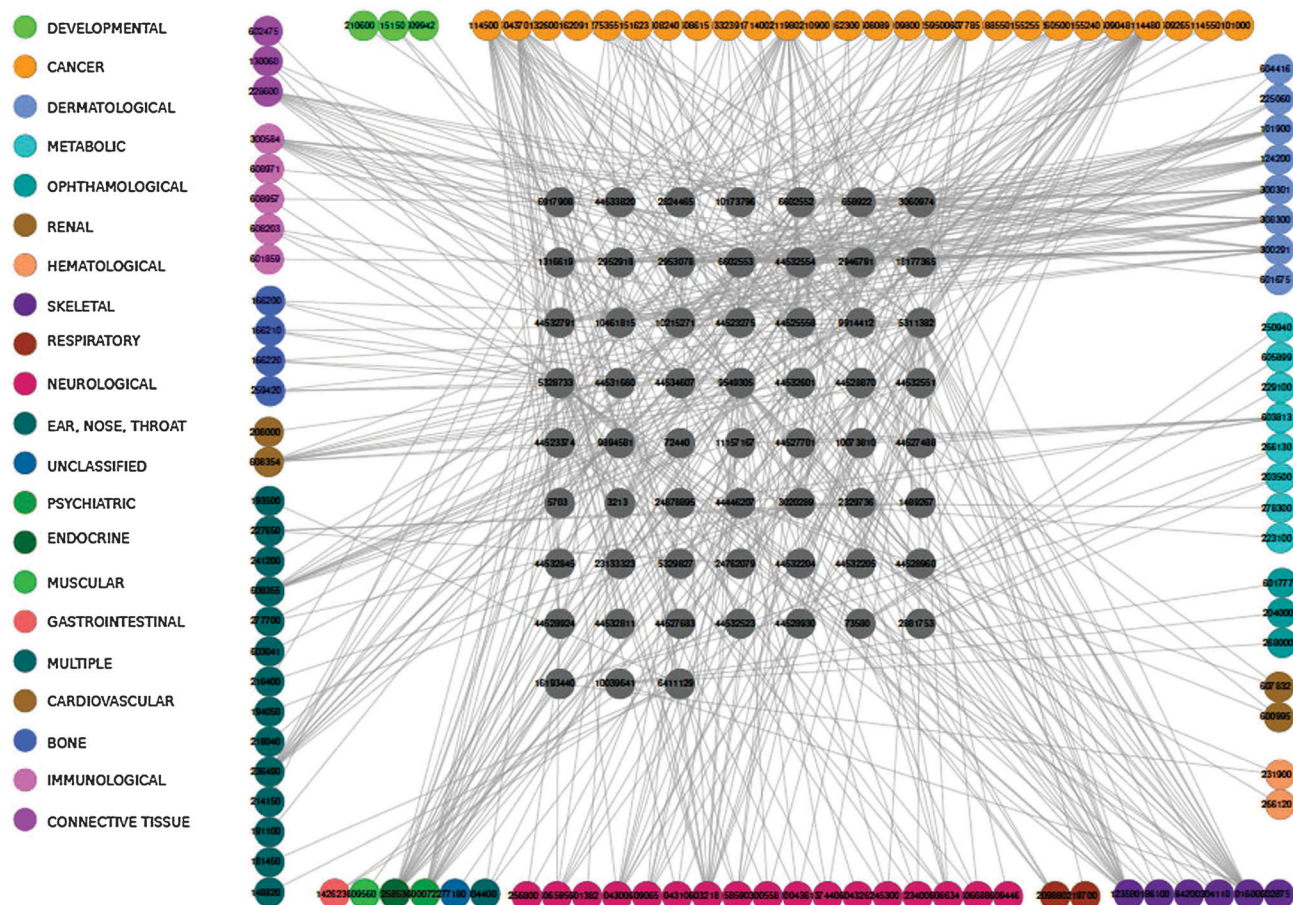


Fig. 5 Disease association network of off-target interactions between GSK compounds and human proteins. Grey nodes: GSK compounds with experimental binding affinity with human proteins, colored nodes: OMIM IDs of targeted human proteins, color key: disorder classes according to Goh *et al.*⁵³

As the available information in CTD in this particular area of the chemical space is very limited, we subsequently explored the effect of the 2512 GSK compounds predicted to be poly-active on humans *via* the integration of chemical–protein, protein–protein and protein–disease associations. We considered only chemical–protein interaction data with direct experimental evidence in ChemProt, while protein complexes and complex–disease associations were retrieved from our in-house human interactome resource.^{29,30} There are in total 59 GSK compounds with experimental activity data for 30 human proteins that form 38 non-redundant protein–protein interaction complexes (see Files S4 and S5 (ESI†) for chemical–protein pairs and protein complexes respectively). Fig. 5 illustrates the disease association network due to known off-target interactions of the selected GSK compounds. Cancer is the disorder class with the most OMIM diseases and the most associations with GSK compounds (26 OMIM diseases and 167 associations in total). The classes of neurological and dermatological disorders follow, with 73 and 64 associations with GSK compounds, respectively.

Methods

The druggable landscape of the *P. falciparum* proteome

The chemical structures of the 13 533 GSK compounds were retrieved from PubChem in SDF format, tagged with a unique

Compound ID (CID). In the cases where two chains were present, only the longest one was kept. MACCS substructure fingerprints were generated with Open Babel (<http://openbabel.org/>). Structural similarity was assessed by the Tanimoto coefficient using a cut-off value of 0.85. Information on chemical–protein interactions was extracted from ChemProt¹⁸ a repository of 700 000 unique chemicals and 30 578 proteins, assembled from both proprietary and freely available databases, including ChEMBL,³¹ BindingDB,³² PDSP Ki Database,³³ DrugBank,³⁴ PharmGKB,³⁵ WOMBAT,³⁶ WOMBAT-PK,³⁶ PubChem bioassay,³⁷ CTD²⁷ and STITCH.³⁸

The protein sequences were retrieved from Batch Entrez (<http://www.ncbi.nlm.nih.gov/sites/batchentrez>) and were blasted against the proteome of *P. falciparum*'s (NCBI Genome Project ID 148) with cut-off *e*-value of 10^{-3} . The genomic location of *P. falciparum* genes was obtained from <http://plasmodb.org/plasmo/>. The preference for chromosomes was evaluated using the Student's *t*-test. GO-terms were assigned to the proteins using Blast2GO (www.blast2go.org/). Heatmaps were created in R, using the *stats* package.³⁹ Hierarchical clustering with Ward's methods and default parameters was applied.

Off-target interactions with the human proteome

Protein–protein interactions were explored through an in-house human interactome network based on experimental data from

humans and 21 model organisms.^{29,30} Using a probabilistic confidence-scoring scheme, all interactions have been validated against a gold standard.⁴⁰ The interactome network is comprised of 22 997 genes, and contains unique protein–protein interactions derived from databases such as BIND,⁴¹ GRID,⁴² MINT,⁴³ dip_full,⁴⁴ HPRD,⁴⁵ intact,⁴⁶ mppi,⁴⁷ MPact,⁴⁸ Reactome⁴⁹ and KEGG.⁵⁰ Data were transferred between organisms by using the Inparanoid orthology database.⁵¹ Protein complexes were subsequently associated to diseases mentioned in OMIM⁵² by using the protein to OMIM mapping displayed in GeneCards (<http://www-bimas.cit.nih.gov/cards/>) database. Because a number of complexes are known to be involved in different diseases, a complex is allowed to be associated with more than one disease. The enrichment of proteins in complexes associated with the same OMIM disease was calculated by using the annotations in GeneCards and the significance of the enrichment was evaluated by using a hypergeometric test. The phenome-interactome resource currently contains 2227 unique disease-related proteins from 1524 complexes, 350 of which are non-redundant. Disorder classes were taken from Goh *et al.*⁵³ and are associated to protein targets *via* their OMIM IDs.

Conclusion

During the last 10–15 years, high-throughput technologies for combinatorial and multi-parallel chemical synthesis and availability of large compound collections that often exceed one million distinct chemical molecules have caused a fundamental change in the drug discovery process conducted in pharmaceutical companies. *In vivo* phenotypic screening of small molecule datasets is seen as a key element for filling the drug discovery pipeline with new chemical compounds. GSK conducted such a massive chemical screen (~2 000 000 chemical entities) to identify growth inhibitors of *P. falciparum*, which yielded 13 533 bioactive molecules. Approximately 8000 of these plasmodial inhibitors were also active against the multi-drug resistant strain. However, the identification of the cellular targets of such bioactive small molecules and the characterization of their modes of action is a major bottleneck in drug discovery.⁵⁴ Identification of the target would facilitate the discovery of drug-cocktails (*e.g.* blocking a whole pathway instead of one protein, or even multiple pathways), while understanding the mode of action would allow the design of efficient drug molecules that can escape drug resistance.⁵⁵

In our study we used the data from the screening of the GSK library of compounds to draft a drug-relevant map of the *P. falciparum* proteome, enriched with functional information. By integrating data from a disease-centric chemical database recently developed in our center (ChemProt¹⁸), we linked ~40% (4645 compounds) of the plasmodial inhibitors with their putative protein targets. The 293 predicted targets contained the six out of the seven verified targets for treating *P. falciparum*-transmitted malaria, while the annotation of the targets revealed established essential functions, *e.g.* nucleotide binding,²⁶ transporter activity,⁵⁶ hypothesized essential functions, *e.g.* protein kinase activity¹⁶ but also novel essential functions, *e.g.* protein binding, transcriptional regulation and protein modification processes. From the 293 predicted targets, we identify

15 proteins as potential Achilles' heel targets in the proteome of *P. falciparum* that fulfill multiple criteria among essentiality for growth, lack of homology with human proteins, lack of resistance emergence and availability of experimental data that support predicted binding to a GSK compound. Proteins PF14_0053, PF10_0289 and PFD0830w appear to be especially interesting on the basis of evading drug resistance.

In the light of multi-target therapy being more effective than single-target therapy when treating infectious diseases, we identified that compounds from the GSK collection, notably CID 23618032, CID 44522264, CID 44531668 and CID 44531763, that are capable of escaping drug resistance, have polypharmacology profiles that cover multiple essential targets and perturb different points of the parasite's metabolic network. Our results could also be used to elucidate drug synergy, and assist in the selection of compounds with synergistic effects that could be used in a combinatorial antimalarial therapy.

However, we should not pass over the fact that both the pharmaceutical industry and the regulatory authorities are concerned by the risk of unexpected side effects of compounds that have been identified as drug leads against parasites based only on high-throughput *in vitro* screening endeavors. In the present study we predict potential off-target effects on the human host with associations to cancer, neurological and dermatological disorders, based on integration of available chemical–protein and protein–protein interaction data.

The development of novel, efficient and inexpensive drugs is an urgent priority. The report of Gamo *et al.*¹⁶ made for the first time accessible to the general research community compounds that were previously proprietary to GSK and offered tremendous opportunities to develop the next generation of antimalarial drugs. We believe that the present work contributes significantly to efforts in identifying the most appropriate targets and the most promising bioactive compounds from the GSK library. Our work suggests that a large number of the *P. falciparum* proteome is potentially druggable and could therefore serve as novel drug targets in the fight against malaria. At the same time, we prioritize compounds from the GSK library in respect to their predicted interactions with the *P. falciparum* proteome and metabolism, with the hope to provide interesting lead compounds to medicinal chemists for further optimization.

Acknowledgements

The authors would like to acknowledge the Innovative Medicines Initiative Joint Undertaking (eTOX). Melani Khodaie is gratefully acknowledged for improving figure quality.

References

- 1 A. G. Nerlich, B. Schraut, S. Dittrich, T. Jelinek and A. R. Zink, *Emerging Infect. Dis.*, 2008, **14**, 1317–1319.
- 2 J. Sachs and P. Malaney, *Nature*, 2002, **415**, 680–685.
- 3 R. T. Eastman and D. A. Fidock, *Nat. Rev. Microbiol.*, 2009, **7**, 864–874.
- 4 R. W. Snow, C. A. Guerra, A. M. Noor, H. Y. Myint and S. I. Hay, *Nature*, 2005, **434**, 214–217.
- 5 T. Anderson, *PLoS Med.*, 2009, **6**, e1000054.
- 6 A. Afonso, P. Hunt, S. Cheesman, A. C. Alves, C. V. Cunha, V. do Rosário and P. Cravo, *Antimicrob. Agents Chemother.*, 2006, **50**, 480–489.

- 7 WHO | World Malaria Report 2009, <http://www.who.int/malaria/publications/atoz/9789241563901/en/index.html>. Last accessed Dec 6 2011. WHO.
- 8 A. M. Dondorp, F. Nosten, P. Yi, D. Das, A. P. Phyto, J. Tarning, K. M. Lwin, F. Ariey, W. Hanpithakpong, S. J. Lee, P. Ringwald, K. Silamut, M. Imwong, K. Chotivanich, P. Lim, T. Herdman, S. S. An, S. Yeung, P. Singhasivanon, N. P. J. Day, N. Lindegardh, D. Socheat and N. J. White, *N. Engl. J. Med.*, 2009, **361**, 455–467.
- 9 S. M. Todryk and A. V. S. Hill, *Nat. Rev. Microbiol.*, 2007, **5**, 487–489.
- 10 I. N. Subbayya, S. S. Ray, P. Balaram and H. Balaram, *Indian J. Med. Res.*, 1997, **106**, 79–94.
- 11 A. S. Tarun, X. Peng, R. F. Dumpit, Y. Ogata, H. Silva-Rivera, N. Camargo, T. M. Daly, L. W. Bergman and S. H. I. Kappe, *Proc. Natl. Acad. Sci. U. S. A.*, 2008, **105**, 305–310.
- 12 M. J. Gardner, N. Hall, E. Fung, O. White, M. Berriman, R. W. Hyman, J. M. Carlton, A. Pain, K. E. Nelson, S. Bowman, I. T. Paulsen, K. James, J. A. Eisen, K. Rutherford, S. L. Salzberg, A. Craig, S. Kyes, M.-S. Chan, V. Nene, S. J. Shallom, B. Suh, J. Peterson, S. Angiuoli, M. Pertea, J. Allen, J. Selengut, D. Haft, M. W. Mather, A. B. Vaidya, D. M. A. Martin, A. H. Fairlamb, M. J. Fraunholz, D. S. Roos, S. A. Ralph, G. I. McFadden, L. M. Cummings, G. M. Subramanian, C. Mungall, J. C. Venter, D. J. Carucci, S. L. Hoffman, C. Newbold, R. W. Davis, C. M. Fraser and B. Barrell, *Nature*, 2002, **419**, 498–511.
- 13 K. G. Le Roch, Y. Zhou, P. L. Blair, M. Grainger, J. K. Moch, J. D. Haynes, P. De La Vega, A. A. Holder, S. Batalov, D. J. Carucci and E. A. Winzler, *Science*, 2003, **301**, 1503–1508.
- 14 TDR Targets. <http://www.tdrtargets.org/>. Last accessed Jan 3 2012.
- 15 W. A. Guiguemde, A. A. Shelat, D. Bouck, S. Duffy, G. J. Crowther, P. H. Davis, D. C. Smithson, M. Connelly, J. Clark, F. Zhu, M. B. Jiménez-Díaz, M. S. Martínez, E. B. Wilson, A. K. Tripathi, J. Gut, E. R. Sharlow, I. Bathurst, F. El Mazouni, J. W. Fowble, I. Forquer, P. L. McGinley, S. Castro, I. Angulo-Barturen, S. Ferrer, P. J. Rosenthal, J. L. Derisi, D. J. Sullivan, J. S. Lazo, D. S. Roos, M. K. Riscoe, M. A. Phillips, P. K. Rathod, W. C. Van Voorhis, V. M. Avery and R. K. Guy, *Nature*, 2010, **465**, 311–315.
- 16 F.-J. Gamo, L. M. Sanz, J. Vidal, C. de Cozar, E. Alvarez, J.-L. Lavandera, D. E. Vanderwall, D. V. S. Green, V. Kumar, S. Hasan, J. R. Brown, C. E. Peishoff, L. R. Cardon and J. F. Garcia-Bustos, *Nature*, 2010, **465**, 305–310.
- 17 S. L. Kinnings, L. Xie, K. H. Fung, R. M. Jackson, L. Xie and P. E. Bourne, *PLoS Comput. Biol.*, 2010, **6**, e1000976.
- 18 O. Taboureaux, S. K. Nielsen, K. Audouze, N. Weinhold, D. Edsgård, F. S. Roque, I. Kouskoumvekaki, A. Bora, R. Curpan, T. S. Jensen, S. Brunak and T. I. Oprea, *Nucleic Acids Res.*, 2011, **39**, D367–D372.
- 19 H. Matter, *J. Med. Chem.*, 1997, **40**, 1219–1229.
- 20 D. Leroy and C. Doerig, *Trends Pharmacol. Sci.*, 2008, **29**, 241–249.
- 21 T. Mitamura and N. M. Q. Palacpac, *Microbes Infect.*, 2003, **5**, 545–552.
- 22 U. Eckstein-Ludwig, R. J. Webb, I. D. A. Van Goethem, J. M. East, A. G. Lee, M. Kimura, P. M. O'Neill, P. G. Bray, S. A. Ward and S. Krishna, *Nature*, 2003, **424**, 957–961.
- 23 P. D. Dobson and D. B. Kell, *Nat. Rev. Drug Discovery*, 2008, **7**, 205–220.
- 24 HomoloGene. Home, <http://www.ncbi.nlm.nih.gov/globalproxy.cvt.dk/homologene/>. Last accessed Jan 3 2012.
- 25 L. He, Y. Wu, L. Lin, J. Wang, Y. Wu, Y. Chen, Z. Yi, M. Liu and X. Pang, *Cancer Sci.*, 2011, **102**, 219–225.
- 26 G. Plata, T.-L. Hsiao, K. L. Olszewski, M. Llinás and D. Vitkup, *Mol. Syst. Biol.*, 2010, **6**, 408.
- 27 A. P. Davis, C. G. Murphy, C. A. Saraceni-Richards, M. C. Rosenstein, T. C. Wieggers and C. J. Mattingly, *Nucleic Acids Res.*, 2009, **37**, D786–D792.
- 28 SIDER, Side Effect Resource.
- 29 K. Lage, N. T. Hansen, E. O. Karlberg, A. C. Eklund, F. S. Roque, P. K. Donahoe, Z. Szallasi, T. S. Jensen and S. Brunak, *Proc. Natl. Acad. Sci. U. S. A.*, 2008, **105**, 20870–20875.
- 30 K. Lage, E. O. Karlberg, Z. M. Störling, P. I. Olason, A. G. Pedersen, O. Rigina, A. M. Hinsby, Z. Tümer, F. Pociot, N. Tommerup, Y. Moreau and S. Brunak, *Nat. Biotechnol.*, 2007, **25**, 309–316.
- 31 P. de Matos, R. Alcantara, A. Dekker, M. Ennis, J. Hastings, K. Haug, I. Spiteri, S. Turner and C. Steinbeck, *Nucleic Acids Res.*, 2009, **38**, D249–D254.
- 32 T. Liu, Y. Lin, X. Wen, R. N. Jorissen and M. K. Gilson, *Nucleic Acids Res.*, 2007, **35**, D198–D201.
- 33 B. Roth, *Pharmacol. Ther.*, 2004, **102**, 99–110.
- 34 D. S. Wishart, C. Knox, A. C. Guo, D. Cheng, S. Shrivastava, D. Tzur, B. Gautam and M. Hassanali, *Nucleic Acids Res.*, 2008, **36**, D901–D906.
- 35 M. Hewett, D. E. Oliver, D. L. Rubin, K. L. Easton, J. M. Stuart, R. B. Altman and T. E. Klein, *Nucleic Acids Res.*, 2002, **30**, 163–165.
- 36 M. Olah, R. Curpan, L. Ostropovici, A. Bora, N. Hadaruga, R. Moldovan, A. Fulas, M. Mracec and T. I. Oprea, in *Chemical Biology: From Small Molecules to Systems Biology and Drug Design*, Wiley-VCH, New York, 2007, pp. 760–786.
- 37 D. L. Wheeler, T. Barrett, D. A. Benson, S. H. Bryant, K. Canese, V. Chetvernin, D. M. Church, M. DiCuccio, R. Edgar, S. Federhen, L. Y. Geer, Y. Kapustin, O. Khovayko, D. Landsman, D. J. Lipman, T. L. Madden, D. R. Maglott, J. Ostell, V. Miller, K. D. Pruitt, G. D. Schuler, E. Sequeira, S. T. Sherry, K. Sirotkin, A. Souvorov, G. Starchenko, R. L. Tatusov, T. A. Tatusova, L. Wagner and E. Yaschenko, *Nucleic Acids Res.*, 2007, **35**, D5–D12.
- 38 M. Kuhn, D. Szklarczyk, A. Franceschini, M. Campillos, C. von Mering, L. J. Jensen, A. Beyer and P. Bork, *Nucleic Acids Res.*, 2009, **38**, D552–D556.
- 39 The R Project for Statistical Computing. <http://www.R-project.org>.
- 40 J. F. Rual, K. Venkatesan, T. Hao, T. Hirozane-Kishikawa, A. Dricot and N. Li, *et al.*, *Nature*, 2005, **437**, 1173–1178.
- 41 G. D. Bader, D. Betel and C. W. Hogue, *Nucleic Acids Res.*, 2003, **31**, 248–250.
- 42 C. Stark, B. J. Breitkreutz, T. Reguly, L. Boucher, A. Breitkreutz and M. Tyers, *Nucleic Acids Res.*, 2006, **34**, D535–D539.
- 43 A. Zanzoni, L. Montecchi-Palazzi, M. Quondam, G. Ausiello, M. Helmer-Citterich and G. Cesareni, *FEBS Lett.*, 2002, **513**, 135–140.
- 44 L. Salwinski, C. S. Miller, A. J. Smith, F. K. Pettit, J. U. Bowie and D. Eisenberg, *Nucleic Acids Res.*, 2004, **32**, D449–D451.
- 45 G. R. Mishra, M. Suresh, K. Kumaran, N. Kannabiran, S. Suresh and P. Bala, *et al.*, *Nucleic Acids Res.*, 2006, **34**, D411–D414.
- 46 H. Hermjakob, L. Montecchi-Palazzi, C. Lewington, S. Mudali, S. Kerrien and S. Orchard, *et al.*, *Nucleic Acids Res.*, 2004, **32**, D452–D455.
- 47 P. Pagel, S. Kovac, M. Oesterheld, B. Brauner, I. Dunger-Kaltenbach and G. Frishman, *et al.*, *Bioinformatics*, 2005, **21**, 832–834.
- 48 U. Guldener, M. Munsterkotter, M. Oesterheld, P. Pagel, A. Ruepp and H. W. Mewes, *et al.*, *Nucleic Acids Res.*, 2006, **34**, D436–D441.
- 49 G. Joshi-Tope, M. Gillespie, I. Vastrik, P. D'Eustachio, E. Schmidt and B. de Bono, *et al.*, *Nucleic Acids Res.*, 2005, **33**, D428–D432.
- 50 M. Kanehisa, S. Goto, M. Hattori, K. F. Aoki-Kinoshita, M. Itoh and S. Kawashima, *et al.*, *Nucleic Acids Res.*, 2006, **34**, D354–D357.
- 51 K. P. O'Brien, M. Remm and E. L. Sonnhammer, *Nucleic Acids Res.*, 2005, **33**, D476.
- 52 A. Hamosh, A. F. Scott, J. S. Amberger, C. A. Bocchini and V. A. McKusick, *Nucleic Acids Res.*, 2005, **33**, D514–D517.
- 53 K.-I. Goh, M. E. Cusick, D. Valle, B. Childs, M. Vidal and A.-L. Barabási, *Proc. Natl. Acad. Sci. U. S. A.*, 2007, **104**, 8685–8690.
- 54 J. P. Overington, B. Al-Lazikani and A. L. Hopkins, *Nat. Rev. Drug Discovery*, 2006, **5**, 993–996.
- 55 T. Shahian, G. M. Lee, A. Lazic, L. A. Arnold, P. Velusamy, C. M. Roels, R. K. Guy and C. S. Craik, *Nat. Chem. Biol.*, 2009, **5**, 640–646.
- 56 H. M. Staines, E. T. Derbyshire, K. Slavic, A. Tattersall, H. Vial and S. Krishna, *Trends Parasitol.*, 2010, **26**, 284–296.